

Generalized Approximate Message-Passing Decoder for Universal Sparse Superposition Codes

Erdem Bıyık*, Jean Barbier† and Mohamad Dia†

* Bilkent University, Ankara, Turkey

† Communication Theory Laboratory, EPFL, Lausanne, Switzerland
erdem.biyik@ug.bilkent.edu.tr, {jean.barbier, mohamad.dia}@epfl.ch

Abstract—Sparse superposition (SS) codes were originally proposed as a capacity-achieving communication scheme over the additive white Gaussian noise channel (AWGNC) [1]. Very recently, it was discovered that these codes are universal, in the sense that they achieve capacity over any memoryless channel under generalized approximate message-passing (GAMP) decoding [2], although this decoder has never been stated for SS codes. In this contribution we introduce the GAMP decoder for SS codes, we confirm empirically the universality of this communication scheme through its study on various channels and we provide the main analysis tools: state evolution and the potential. We also compare the performance of GAMP with the Bayes-optimal MMSE decoder. We empirically illustrate that despite the presence of a phase transition preventing GAMP to reach the optimal performance, spatial coupling allows to boost the performance that eventually tends to capacity in a proper limit. We also prove that, in contrast with the AWGNC case, SS codes for binary input channels have a vanishing error floor in the limit of large codewords. Moreover, the performance of Hadamard-based encoders is assessed for practical implementations.

I. INTRODUCTION

Sparse superposition codes were introduced by Barron and Joseph for communication over the AWGNC [1]. These codes were proven to achieve the Shannon capacity using power allocation and various efficient decoders [3, 4]. A decoder based on approximate message-passing (AMP), originally developed for compressed sensing [5, 6], was introduced in [7]. The authors in [8] proved, using the state evolution (SE) analysis [9, 10], that AMP allows to achieve capacity using power allocation. At the same time, spatially coupled SS codes were introduced in [11, 12] and empirically shown to approach capacity under AMP without power allocation and to perform much better than power allocated ones. Recently, AMP for spatially coupled SS codes was shown to saturate the so-called *potential threshold*, related to the Bayes-optimal MMSE performance, which tends to capacity in a proper limit [13].

This set of works combined with the excellent performance of SS codes over the AWGNC motivated their study for any memoryless channel under GAMP decoding [2]. GAMP was introduced as a generalization of AMP for generalized estimation [14]. In [2] the authors showed that, under the assumption that SE [10] tracks GAMP for SS codes, spatially coupled SS codes achieve the capacity of any memoryless channel under GAMP decoding. However, GAMP has never been explicitly stated or tested as a decoder for SS codes other than for the AWGNC, in which case GAMP and AMP are identical. In

this work we fill this gap by studying the GAMP decoder for SS codes over various memoryless channels. We focus on the AWGNC (for completeness with previous studies [11, 12]), binary erasure channel (BEC), Z channel (ZC) and binary symmetric channel (BSC). However, the present decoder and analysis remain valid for any memoryless channel.

Our experiments confirm that SE recursion of [2] accurately tracks GAMP. Using the potential of the code we also compare the performance of GAMP to the optimal MMSE decoder. In addition, our empirical study confirms the asymptotic (in the block-length) results of [2]: the performance of SS codes under GAMP decoding can be significantly increased towards capacity using spatial coupling, as already observed for the AWGNC [12]. Moreover, we prove that for binary input channels, SS codes have a vanishing error floor in the limit of large codewords even with finite sparsity. This implies that when GAMP decoding is possible, it is asymptotically perfect until some threshold, a very promising feature which is *not* present for the real-valued input AWGNC. Keeping in mind practicality, we focus our empirical study on Hadamard-based coding operators that allow to drastically reduce the encoding and decoding complexity, while maintaining good performance for moderate block-lengths [11].

II. SPARSE SUPERPOSITION CODES: SETTING

In SS codes, the *message* $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_L]$ is a vector made of L B -dimensional *sections*. Each section \mathbf{x}_l , $l \in \{1, \dots, L\}$, satisfies a hard constraint: it has a single non-zero component equals 1 whose position encodes the symbol to transmit. $B \geq 2$ is the *section size* (or alphabet size) and we set $N := LB$. For the theoretical analysis we consider random codes generated by a *coding matrix* $\mathbf{A} \in \mathbb{R}^{M \times N}$ drawn from the ensemble of Gaussian matrices with i.i.d entries $\sim \mathcal{N}(0, \sigma_A^2)$. For the practical implementation, fast Hadamard-based operators are used instead as they exhibit very good performances. Despite the lack of rigor in the analysis for such operators, they remain good predictive tools [11]. The *codeword* is $\mathbf{Ax} \in \mathbb{R}^M$. We enforce the power constraint $\mathbb{E}[\|\mathbf{Ax}\|_2^2]/M = 1$ by tuning σ_A^2 . The cardinality of the code is B^L . Hence, the (design) rate is $R = L \log_2(B)/M$ and the code is specified by (M, R, B) .

The aim is to communicate through a known memoryless channel W . This requires to map the continuous-valued codeword onto the input alphabet of W . The concatenation of this mapping operation and the channel it-

Algorithm 1 GAMP ($\mathbf{y}, \mathbf{A}, B, t_{\max}, u$)

```

1:  $\hat{\mathbf{x}}(0) = \mathbf{0}_{N,1}$ ,  $\boldsymbol{\tau}^x(0) = (1/B)\mathbf{1}_{N,1}$ 
2:  $\mathbf{s}_{(-1)} = \mathbf{0}_{M,1}$ ,  $t = 0$ ,  $e_{(0)} = \infty$  ▷ Initializations
3: while  $t \leq t_{\max}$  and  $e_{(t)} \geq u$  do
4:    $\boldsymbol{\tau}_{(t)}^p = \mathbf{A}^{\circ 2} \boldsymbol{\tau}_{(t)}^x$ 
5:    $\mathbf{p}_{(t)} = \mathbf{A} \hat{\mathbf{x}}_{(t)} - \boldsymbol{\tau}_{(t)}^p \circ \mathbf{s}_{(t-1)}$  ▷ Output linear step
6:    $\boldsymbol{\tau}_{(t)}^s = -g'_{\text{out}}(\mathbf{p}_{(t)}, \mathbf{y}, \boldsymbol{\tau}_{(t)}^p)$ 
7:    $\mathbf{s}_{(t)} = g_{\text{out}}(\mathbf{p}_{(t)}, \mathbf{y}, \boldsymbol{\tau}_{(t)}^p)$  ▷ Output non-linear step
8:    $\boldsymbol{\tau}_{(t)}^r = (((\boldsymbol{\tau}_{(t)}^s)^\top \mathbf{A}^{\circ 2} \boldsymbol{\tau}_{(t)}^p)^\circ)^{-1}$ 
9:    $\mathbf{r}_{(t)} = \hat{\mathbf{x}}_{(t)} + \boldsymbol{\tau}_{(t)}^r \circ (\mathbf{s}_{(t)}^\top \mathbf{A})^\top$  ▷ Input linear step
10:   $\boldsymbol{\tau}_{(t+1)}^x = \boldsymbol{\tau}_{(t)}^r \circ g'_{\text{in}}(\mathbf{r}_{(t)}, \boldsymbol{\tau}_{(t)}^r)$ 
11:   $\hat{\mathbf{x}}_{(t+1)} = g_{\text{in}}(\mathbf{r}_{(t)}, \boldsymbol{\tau}_{(t)}^r)$  ▷ Input non-linear step
12:   $e_{(t)} = \|\hat{\mathbf{x}}_{(t+1)} - \hat{\mathbf{x}}_{(t)}\|_2^2 / L$ 
13:   $t = t + 1$ 
14: return  $\hat{\mathbf{x}}_{(t)}$  ▷ The prediction scores for each bit

```

self can be interpreted as an *effective memoryless channel* $P_{\text{out}}(\mathbf{y}|\mathbf{A}\mathbf{x}) = \prod_{\mu=1}^M P_{\text{out}}(y_\mu|[\mathbf{A}\mathbf{x}]_\mu)$. For the channels we focus on, $P_{\text{out}}(y_\mu|[\mathbf{A}\mathbf{x}]_\mu)$ is expressed as follows:

- AWGNC: $\mathcal{N}(y_\mu|[\mathbf{A}\mathbf{x}]_\mu, 1/\text{snr})$,
- BEC: $(1-\epsilon)\delta(y_\mu - \text{sign}([\mathbf{A}\mathbf{x}]_\mu)) + \epsilon\delta(y_\mu)$,
- BSC: $(1-\epsilon)\delta(y_\mu - \text{sign}([\mathbf{A}\mathbf{x}]_\mu)) + \epsilon\delta(y_\mu + \text{sign}([\mathbf{A}\mathbf{x}]_\mu))$,
- ZC: $\delta(\text{sign}([\mathbf{A}\mathbf{x}]_\mu) + 1)(\epsilon\delta(y_\mu - 1) + (1-\epsilon)\delta(y_\mu + 1)) + \delta(\text{sign}([\mathbf{A}\mathbf{x}]_\mu) - 1)\delta(y_\mu - 1)$,

where snr is the signal-to-noise ratio of the AWGNC, ϵ the erasure or flip probability of the BEC, ZC and BSC. The sign maps the Gaussian distributed codeword components onto the input alphabets of the binary input channels.

Note that for the asymmetric ZC, the symmetric map $\text{sign}([\mathbf{A}\mathbf{x}]_\mu)$ leads to a sub-optimal uniform input distribution. The *symmetric capacity* of the ZC differs from Shannon's capacity but the difference is small, and similarly for the algorithmic threshold, see [2]. We thus consider this symmetric setting for the sake of simplicity. The other channels are symmetric, this map thus leads to the optimal input distribution.

III. THE GAMP DECODER

We consider a Bayesian setting and associate to the message the posterior $P(\mathbf{x}|\mathbf{y}, \mathbf{A}) = P_{\text{out}}(\mathbf{y}|\mathbf{A}\mathbf{x})P_0(\mathbf{x})/P(\mathbf{y}|\mathbf{A})$. The hard constraints for the sections are enforced by the prior $P_0(\mathbf{x}) = \prod_{l=1}^L p_0(\mathbf{x}_l)$ with $p_0(\mathbf{x}_l) = B^{-1} \sum_{i \in l} \delta_{x_i, 1} \prod_{j \in l, j \neq i} \delta_{x_j, 0}$, where $\{i \in l\}$ are the B scalar components indices of the section l . The GAMP decoder aims at performing MMSE estimation by approximating the posterior mean of each section.

In the GAMP decoder Algorithm 1, \circ denotes element-wise operations. GAMP was originally derived for *scalar* estimation. In this generalization to the vectorial setting of SS codes, whose derivation is similar to the one of AMP for SS codes found in [12], only the input non-linear steps differ from canonical GAMP [14]: here the so-called *denoiser* g_{in} acts *sectionwise* instead of componentwise. In full generality, it is defined as [14] $g_{\text{in}}(\mathbf{r}, \boldsymbol{\tau}) := \mathbb{E}[\mathbf{X}|\mathbf{R} = \mathbf{r}]$ for the random variable $\mathbf{R} = \mathbf{X} + \hat{\mathbf{Z}}$ with $\mathbf{X} \sim P_0$ and $\hat{\mathbf{Z}} \sim \mathcal{N}(0, \text{diag}(\boldsymbol{\tau}))$.

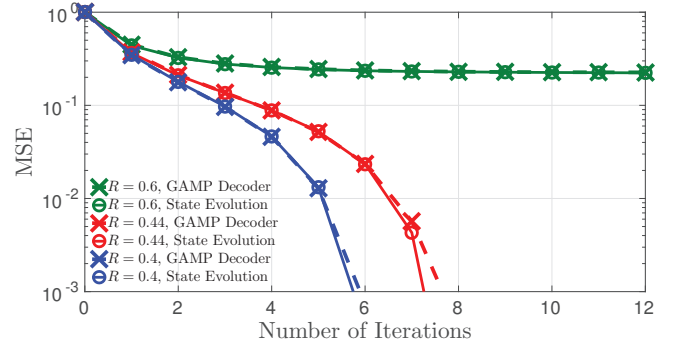


Fig. 1: SE tracking the GAMP decoder (averaged over 100 random instances) over the BEC with erasure probability $\epsilon = 0.1$ and for $L = 2^{11}$, $B = 4$ and Gaussian coding operators. Monte carlo integration with 2×10^4 samples is used for the computation of SE. The algorithmic threshold $R_{\text{GAMP}} \approx 0.55$ and the green curves are for a rate above it: decoding fails. In contrast, the blue and red curves are below R_{GAMP} : decoding succeeds. After the last points of these curves, both SE and the GAMP curves fall to 0 MSE.

Moreover, the estimate of the posterior variance, which quantifies how “confident” GAMP is in its current estimate, equals $\boldsymbol{\tau} \circ g'_{\text{in}}(\mathbf{r}, \boldsymbol{\tau}) = \mathbb{E}[\mathbf{X}^{\circ 2}|\mathbf{R} = \mathbf{r}] - g_{\text{in}}(\mathbf{r}, \boldsymbol{\tau})^{\circ 2}$ (g'_{in} is the componentwise partial derivative w.r.t its first argument, and similarly for g'_{out}). Plugging P_0 yields the componentwise expression of the denoiser and the variance term:

$$\begin{cases} [g_{\text{in}}(\mathbf{r}, \boldsymbol{\tau})]_i &= \frac{\exp((2r_i - 1)/(2\tau_i))}{\sum_{j \in l_i} \exp((2r_j - 1)/(2\tau_j))}, \\ [\boldsymbol{\tau} \circ g'_{\text{in}}(\mathbf{r}, \boldsymbol{\tau})]_i &= [g_{\text{in}}(\mathbf{r}, \boldsymbol{\tau})]_i (1 - [g_{\text{in}}(\mathbf{r}, \boldsymbol{\tau})]_i), \end{cases}$$

l_i being the section to which belong the i^{th} scalar component.

In contrast with g_{in} that only depends on P_0 , g_{out} depends on the communication channel and acts componentwise. Its general form and specific expressions for the studied channels are given in Table I along with the necessary derivatives.

The complexity of GAMP is dominated by the $\mathcal{O}(MN) = \mathcal{O}(L^2 B \ln(B))$ matrix-vector multiplications. In terms of memory, it is necessary to store \mathbf{A} which can be problematic for large codes. Fast Hadamard-based operators constructed as in [11], with random sub-sampled modes of the full Hadamard operator, allow to achieve a lower $\mathcal{O}(L \ln(B) \ln(BL))$ decoding complexity and strongly reduce the memory need [12, 15].

IV. STATE EVOLUTION AND THE POTENTIAL

We now present the analysis tools of the $L \rightarrow \infty$ performance of SS codes under GAMP and MMSE decoding when Gaussian matrices are used: state evolution and the potential.

A. State evolution

The asymptotic performance of GAMP (asymptotic in L, M for any fixed R and $B \geq 2$) with Gaussian i.i.d coding matrices is tracked by SE, a scalar recursion [2, 9, 10, 14] analogous to density evolution for low-density parity-check codes. Note that although SE is not rigorous for the present vectorial setting, the rigorous analysis of [8] and the present empirical results strongly suggest that *it is exact*.

The aim is to compute the asymptotic MSE of the GAMP estimate $E_{(t)} := \lim_{L \rightarrow \infty} \|\hat{\mathbf{x}}_{(t)} - \mathbf{x}\|_2^2 / L$. It turns out that this

TABLE I: The expressions for g_{out} , $-g'_{\text{out}}$ and \mathcal{F} .

	$[g_{\text{out}}(\mathbf{p}, \mathbf{y}, \boldsymbol{\tau})]_i$	$[-g'_{\text{out}}(\mathbf{p}, \mathbf{y}, \boldsymbol{\tau})]_i$	$\mathcal{F}(p E)$
General	$\frac{(\mathbb{E}[Z_i p_i, y_i, \tau_i] - p_i)/\tau_i}{Y_i \sim P_{\text{out}}(\cdot z_i), Z_i \sim \mathcal{N}(p_i, \tau_i)}$	$\frac{(\tau_i - \text{Var}[Z_i p_i, y_i, \tau_i])/\tau_i^2}{Y_i \sim P_{\text{out}}(\cdot z_i), Z_i \sim \mathcal{N}(p_i, \tau_i)}$	See (1)
AWGNC	$\frac{y_i - p_i}{\tau_i + 1/\text{snr}}$	$\frac{1}{\tau_i + 1/\text{snr}}$	$\frac{1}{1/\text{snr} + E}$
BEC	$\frac{(p_i - k_i)h_i^+ + (p_i + k_i)h_i^- + 2\epsilon\delta(y_i)p_i}{Z_{\text{BEC}}\tau_i} - \frac{p_i}{\tau_i}$	$\frac{1}{\tau_i} - \frac{(p_i^2 + \tau_i - k_i')h_i^+ + (p_i^2 + \tau_i + k_i')h_i^- + 2\epsilon\delta(y_i)(\tau_i + p_i^2)}{Z_{\text{BEC}}\tau_i^2} + ([g_{\text{out}}(\mathbf{p}, \mathbf{y}, \boldsymbol{\tau})]_i + \frac{p_i}{\tau_i})^2$	$\frac{Q'^2(1-\epsilon)}{Q(1-Q)}$
ZC	$\frac{(p_i - k_i)v_i^+ + (p_i + k_i)\delta(y_i - 1)}{Z_{\text{ZC}}\tau_i} - \frac{p_i}{\tau_i}$	$\frac{1}{\tau_i} - \frac{(p_i^2 + \tau_i - k_i')v_i^+ + (p_i^2 + \tau_i + k_i')\delta(y_i - 1)}{Z_{\text{ZC}}\tau_i^2} + ([g_{\text{out}}(\mathbf{p}, \mathbf{y}, \boldsymbol{\tau})]_i + \frac{p_i}{\tau_i})^2$	$\frac{Q'^2(1-\epsilon)^2}{Q + \epsilon(1-Q)} + \frac{Q'^2(1-\epsilon)}{1-Q}$
BSC	$\frac{(p_i - k_i)v_i^+ + (p_i + k_i)v_i^-}{Z_{\text{BSC}}\tau_i} - \frac{p_i}{\tau_i}$	$\frac{1}{\tau_i} - \frac{(p_i^2 + \tau_i - k_i')v_i^+ + (p_i^2 + \tau_i + k_i')v_i^-}{Z_{\text{BSC}}\tau_i^2} + ([g_{\text{out}}(\mathbf{p}, \mathbf{y}, \boldsymbol{\tau})]_i + \frac{p_i}{\tau_i})^2$	$\frac{Q'^2(1-2\epsilon)^2}{(Q + \epsilon - 2\epsilon Q)(1 - Q - \epsilon + 2\epsilon Q)}$
$h_i^+ = (1-\epsilon)\delta(y_i+1), \quad h_i^- = (1-\epsilon)\delta(y_i-1), \quad v_i^+ = (1-\epsilon)\delta(y_i+1) + \epsilon\delta(y_i-1), \quad v_i^- = (1-\epsilon)\delta(y_i-1) + \epsilon\delta(y_i+1),$ $k_i = \exp(-\frac{p_i^2}{2\tau_i})\sqrt{2\tau_i/\pi}\text{erf}(\frac{p_i}{\sqrt{2\tau_i}})p_i, \quad k_i' = k_i p_i + \text{erf}(\frac{p_i}{\sqrt{2\tau_i}})\tau_i, \quad Q = \frac{1}{2}\text{erfc}(\frac{\sqrt{p}}{\sqrt{2E}}), \quad Q' = \exp(-\frac{p^2}{2E})/\sqrt{2\pi E}$ $Z_{\text{BEC}} = \text{erfc}(\frac{p_i}{\sqrt{2\tau_i}})h_i^+ + (1 + \text{erf}(\frac{p_i}{\sqrt{2\tau_i}}))h_i^- + 2\epsilon\delta(y_i), \quad Z_{\text{ZC}} = \text{erfc}(\frac{p_i}{\sqrt{2\tau_i}})v_i^+ + (1 + \text{erf}(\frac{p_i}{\sqrt{2\tau_i}}))\delta(y_i - 1), \quad Z_{\text{BSC}} = \text{erfc}(\frac{p_i}{\sqrt{2\tau_i}})v_i^+ + (1 + \text{erf}(\frac{p_i}{\sqrt{2\tau_i}}))v_i^-$			

is equivalent to recursively compute the MMSE $T(E) := \mathbb{E}_{\mathbf{S}, \mathbf{Z}}[\|\mathbf{S} - \mathbb{E}[\mathbf{X}|\mathbf{S} + (\Sigma(E)/b)\mathbf{Z}]\|_2^2]$ of a single section ($\mathbf{S} \sim p_0$) sent through an equivalent AWGNC ($\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_B)$) of noise variance $(\Sigma(E)/b)^2$, $b^2 := \log_2(B)$. This formulation is valid for any memoryless channel [2], P_{out} being reflected in

$$\begin{cases} \Sigma(E) &:= \sqrt{R}[\int dp \mathcal{N}(p|0, 1-E)\mathcal{F}(p|E)]^{-1/2}, \\ \mathcal{F}(p|E) &:= \int dy f(y|p, E)(\partial_x \ln f(y|x, E))_{x=p}^2, \\ f(y|p, E) &:= \int dz P_{\text{out}}(y|z)\mathcal{N}(z|p, E). \end{cases} \quad (1)$$

\mathcal{F} is the Fisher information of p associated with f , see Table I. The p integral in Σ can be numerically computed for the BEC, ZC and BSC. Define

$$\begin{cases} g_{\text{in}}^{(1)}(\Sigma, \mathbf{z}) &:= [1 + e^{-\frac{b^2}{2\Sigma}} \sum_{j=2}^B e^{\frac{b}{2\Sigma}(z_j - z_1)}]^{-1}, \\ g_{\text{in}}^{(2)}(\Sigma, \mathbf{z}) &:= [1 + e^{\frac{b^2}{2\Sigma} + (z_1 - z_2)\frac{b}{2\Sigma}} + \sum_{k=3}^B e^{(z_k - z_2)\frac{b}{2\Sigma}}]^{-1}. \end{cases}$$

The MMSE of the equivalent AWGNC is obtained after simple algebra [12] and reads

$$T(E) = \mathbb{E}_{\mathbf{Z}}[(g_{\text{in}}^{(1)}(\Sigma(E), \mathbf{Z}) - 1)^2 + (B-1)g_{\text{in}}^{(2)}(\Sigma(E), \mathbf{Z})^2].$$

Here $g_{\text{in}}^{(1)}$ is interpreted as the posterior mean approximated by GAMP of the non-zero component in the transmitted section while $g_{\text{in}}^{(2)}$ corresponds to the remaining components. The SE recursion tracking the MSE of GAMP is then

$$E_{(t+1)} = T(E_{(t)}), \quad t \geq 0, \quad (2)$$

initialized with $E_{(0)} = 1$. Hence, the asymptotic MSE reached by GAMP upon convergence is $E_{(\infty)}$. Moreover, define the asymptotic error floor of SS codes E_* as the fixed point of SE (2) initialized from $E_{(0)} = 0$. Fig. 1 shows that SE tracks GAMP on the BEC. Note that the section error rate (SER) of GAMP, the fraction of wrongly decoded sections after hard thresholding of $\hat{\mathbf{x}}_{(t)}$, can be asymptotically tracked through a simple one-to-one mapping from $E_{(t)}$ [7, 12].

Under GAMP decoding SS codes exhibit, as $L \rightarrow \infty$, a sharp phase transition at an *algorithmic threshold* R_{GAMP} below Shannon's capacity. R_{GAMP} is defined as the highest rate such that for $R \leq R_{\text{GAMP}}$, (2) has a unique fixed point $E_{(\infty)} = E_*$ (see [2] for formal definitions). In this regime GAMP decodes well, see red and blue curves of Fig. 1. If $R > R_{\text{GAMP}}$ GAMP decoding fails, see green curve. As we will see in the next

sections, spatial coupling may allow to boost the performance of the scheme by increasing the GAMP algorithmic threshold.

B. Potential formulation

The SE (2) is associated with a *potential* $F_u(E)$, whose *stationary points correspond to the fixed points of SE*: $\partial_E F_u(E)|_{E_0} = 0 \Leftrightarrow T(E_0) = E_0$. For SS codes it is [2]:

$$\begin{cases} F_u(E) &:= U_u(E) - S_u(\Sigma(E)) \\ U_u(E) &:= -\frac{E}{2\ln(2)\Sigma(E)^2} - \frac{1}{R}\mathbb{E}_{\mathbf{Z}}[\int dy \phi \log_2(\phi)], \\ S_u(\Sigma(E)) &:= \mathbb{E}_{\mathbf{Z}}[\log_B(1 + \sum_{i=2}^B e_i(\mathbf{Z}, \Sigma(E)/b))], \end{cases}$$

where $\phi = \phi(y|Z, E) := \int ds P_{\text{out}}(y|s)\mathcal{N}(s|Z\sqrt{1-E}, E)$, $Z \sim \mathcal{N}(0, 1)$ and $e_i(\mathbf{Z}, x) := \exp((Z_i - Z_1)/x - 1/x^2)$.

It has been recently shown for random linear estimation, including compressed sensing and SS codes with AWGN [16, 17], that $\min_{E \in [0, 1]} F_u(E)$ equals the asymptotic mutual information (up to a trivial additive term) and that $\tilde{E} := \arg\min_{E \in [0, 1]} F_u(E)$ equals the asymptotic MMSE. A proof for all memoryless channels remains to be done, but we conjecture that it remains true under mild conditions on P_{out} .

Using these properties of the potential and its link with the fixed points of SE, it is possible to assess the performances of the GAMP and MMSE decoders by looking at its minima. GAMP decoding is possible (and asymptotically optimal as it reaches the MMSE \tilde{E} , black dot in Fig. 2) for rates lower or equal to R_{GAMP} , whose equivalent definition is the smallest solution of $\partial F_u/\partial E = \partial^2 F_u/\partial E^2 = 0$; in other words it is the smallest rate at which a horizontal inflection point appears in the potential, see blue and red curves in Fig. 2. For $R \in]R_{\text{GAMP}}, R_{\text{pot}}[$, referred to as the *hard phase*, the potential possesses another local minimum (red dot) and the corresponding “bad” fixed point of SE prevents GAMP to reach \tilde{E} ; hence decoding fails (yellow curves). Finally, the rate at which the local and global minima switch roles is the *potential threshold* R_{pot} (purple curves). Optimal decoding is possible as long as $R < R_{\text{pot}}$ as the MMSE switches at R_{pot} from a “low” to a “high” value. At higher rates GAMP is again optimal but leads to poor results as decoding is impossible. Note that if $R < R_{\text{pot}}$, then $E_* = \tilde{E}$.

In the hard phase, where two minima coexist, spatial coupling enables decoding [11] by “effectively suppressing” the

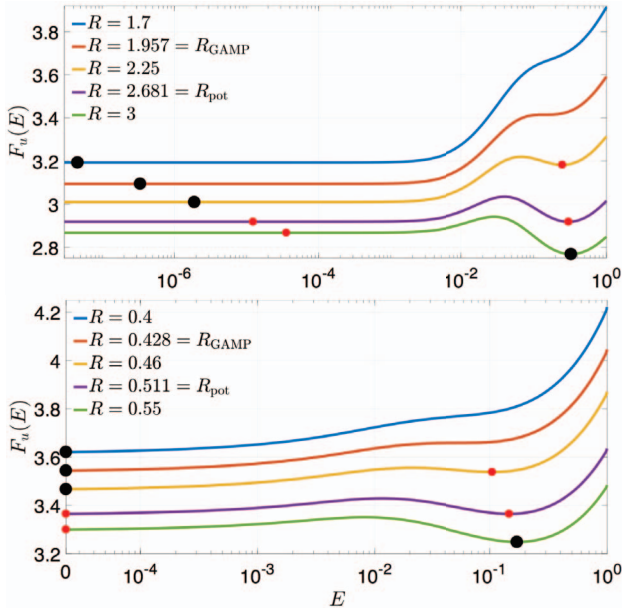


Fig. 2: Potential for the AWGNC with $\text{snr}=100$ (top) and the BEC with $\epsilon = 0.1$ (bottom), in both cases with $B = 2$. The MMSE is the $\text{argmin } F_u(E)$ (black dot). When the min. is unique (i.e. $R < R_{\text{GAMP}}$, blue curve) or if the global min. is the rightmost one ($R > R_{\text{pot}}$, green curve), GAMP is asymptotically optimal, despite that if $R > R_{\text{pot}}$ it leads to poor results. The red dot is the local min., preventing GAMP to decode if $R \in]R_{\text{GAMP}}, R_{\text{pot}}[$ (yellow curve).

spurious local minimum of the potential. It implies that the *algorithmic threshold of spatially coupled SS codes* R_{GAMP}^c , the highest attainable rate using coupled codes under GAMP decoding [2], saturates the potential threshold R_{pot} in the limit of infinite coupled chains. This phenomenon is referred to as *threshold saturation* and is understood as the generic mechanism behind the excellent performances of coupled codes [2, 18]. Moreover, a very interesting feature of SS codes is that R_{pot} itself approaches the capacity as $B \rightarrow \infty$ [2]. These phenomena imply together that in these limits (infinite chain length and B), spatially coupled SS codes under GAMP decoding are *universal* in the sense that they achieve the Shannon capacity of all memoryless channels.

C. Vanishing error floor for binary input memoryless channels

Another promising feature of SS codes is related to their error floor. In the real-valued input AWGNC case, an error floor always exists but it can be made arbitrary small by increasing B [2, 12]. In contrast, in the BEC, ZC and BSC cases (more generally for binary input memoryless channels), we now prove that *as $L \rightarrow \infty$ the error floor vanishes for any ϵ and B* . This implies that when $E_* = \tilde{E}$ optimal decoding is asymptotically perfect, and thus GAMP decoding as well for $R \leq R_{\text{GAMP}}$. This is actually verified in practice for GAMP where perfect decoding is statistically possible even for moderate block-lengths, see blue and red curves of Fig. 1.

The proof of $E_* = 0$, i.e the existence of the *trivial fixed point* $T(0) = 0$ of (2), does not guarantee that this is the global minimum of the potential in the hard phase; i.e it is *a priori* possible that $E_* \neq \tilde{E}$. Nevertheless, our careful numerical work

indicates that *there exist at most two fixed points of SE at the same time or equivalently two minima in the potential*, namely $E_* = \tilde{E} \neq E_{(\infty)}$ if $R \in]R_{\text{GAMP}}, R_{\text{pot}}[$ or $E_* \neq \tilde{E} = E_{(\infty)}$ if $R > R_{\text{pot}}$ (at least for the studied cases), see Fig. 2. This also agrees with the $B \rightarrow \infty$ analysis of the potential [2, 12].

Let us now prove that $E_* = 0$ for the BEC, the proof for other binary input channels being similar. It starts by noticing, from the definition of $T(E)$ as the MMSE of an AWGNC with noise parameter $\Sigma(E)$, that a sufficient condition for $T(0) = 0$ is $\lim_{E \rightarrow 0} \Sigma(E) = 0$; indeed no noise implies vanishing MMSE. From (1) this condition is equivalent to $\lim_{E \rightarrow 0} I_{\mathbb{R}}(E) = \infty$ that we now prove, where $I_A(E) := \int_{\mathcal{A}} dp \mathcal{N}(p|0, 1-E) \mathcal{F}(p|E)$. Consider instead $I_{\mathcal{E}}(E)$ where $\mathcal{E} := [E - \sqrt{E}, E + \sqrt{E}]$. Using Table I for the expression of $\mathcal{F}(p|E)$ for the BEC, this restricted integral is

$$I_{\mathcal{E}}(E) = \frac{(1-\epsilon)(2\pi)^{-3/2}}{E\sqrt{1-E}} \int_{\mathcal{E}} dp \frac{e^{-\frac{p^2}{2(1-E)} - \frac{p^2}{E}}}{Q(p, E)(1-Q(p, E))}.$$

Here $Q(p, E) \in [C_E, 1 - C_E]$, with $\lim_{E \rightarrow 0} C_E > 0$ for $p \in \mathcal{E}$, $E \leq 1$. This implies that $K(E) := \max_{p \in \mathcal{E}} Q(p, E)(1 - Q(p, E)) = \mathcal{O}(1)$. Since the interval \mathcal{E} is of size $2\sqrt{E}$, then

$$I_{\mathcal{E}}(E) \geq \frac{(1-\epsilon)(2\pi)^{-3/2} 2\sqrt{E}}{E\sqrt{1-E}K(E)} e^{-\frac{(E+\sqrt{E})^2}{2(1-E)} - \frac{(E+\sqrt{E})^2}{E}}. \quad (3)$$

From this we can assert that $\lim_{E \rightarrow 0} I_{\mathcal{E}}(E) = \infty$. Moreover $I_{\mathcal{E}}(E) < I_{\mathbb{R}}(E)$ as $\mathcal{F}(p|E) \geq 0$ (recall it is a Fisher information) and thus $\lim_{E \rightarrow 0} I_{\mathbb{R}}(E) = \infty$ which ends the proof.

For the BSC and ZC the proof is similar, the main ingredient being the squared Gaussian Q^2 at the numerator of $\mathcal{F}(p|E)$, see Table I, which leads to similar expressions as (3) and thus the $1/\sqrt{E}$ divergence when $E \rightarrow 0$. We believe that the same mechanism holds for any binary input memoryless channel, implying a vanishing error floor as well as asymptotic perfect decoding of GAMP below the algorithmic threshold.

V. NUMERICAL EXPERIMENTS

In Fig. 3 we compare the optimal and GAMP performances in terms of attainable rate, denoted by R_{pot} and R_{GAMP} respectively. For all channels, there exists, as long as the noise is not “too high”, a hard phase where GAMP is sub-optimal. Moreover, the use of Hadamard-based operators have a performance cost w.r.t Gaussian ones but which vanishes as B increases; they both have the same algorithmic threshold for B large enough (but still practical, $B \geq 64$ being enough).

Consider Gaussian matrices. An interesting feature is that in contrast with the AWGNC case [12], R_{GAMP} for these binary input channels is not monotonously decreasing in B ; it increases until some B (that may be large) but, although it may be hard to observe numerically (except for BEC), it then decreases to reach $\lim_{B \rightarrow \infty} R_{\text{GAMP}} = \mathcal{F}(0|1)/(2 \ln(2)) < C$ [2]. However, a gap to capacity C persists as long as spatial coupling is not employed.

Spatial coupling allows important improvement towards R_{pot} even in practical settings, confirming the universality of coupled SS codes under GAMP decoding as we have

$\lim_{B \rightarrow \infty} R_{\text{pot}} = C$ [2]. The mismatch between R_{pot} (blue) and R_{GAMP}^c (green) is due to finite size effects which are more evident in coupled codes (both chain lengths and coupling windows should go to infinity after L for R_{GAMP}^c to saturate R_{pot}). A Matlab demo for reproducing these results is available: <http://github.com/erdem-biyik/gamp-for-sparse-superposition>

ACKNOWLEDGMENTS

We thank Nicolas Macris, Florent Krzakala and Rüdiger Urbanke for comments as well as Alper Kose and Berke Aral Sonmez for an early stage study of GAMP. J.B and M.D acknowledge funding by the SNSF grant no. 200021-156672.

REFERENCES

- [1] A. Barron and A. Joseph, "Toward fast reliable communication at rates near capacity with gaussian noise," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, June 2010, pp. 315–319.
- [2] J. Barbier, M. Dia, and N. Macris, "Threshold Saturation of Spatially Coupled Sparse Superposition Codes for All Memoryless Channels," in *Information Theory workshop (ITW), 2016 IEEE*, Sep. 2016.
- [3] A. Joseph and A. R. Barron, "Fast sparse superposition codes have near exponential error probability for $R < C$," *IEEE Tans. on Information Theory*, vol. 60, no. 2, pp. 919–942, 2014.
- [4] A. R. Barron and S. Cho, "High-rate sparse superposition codes with iteratively optimal estimates," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 120–124.
- [5] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009.
- [6] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, vol. P08009, 2012.
- [7] J. Barbier and F. Krzakala, "Replica analysis and approximate message passing decoder for superposition codes," in *Information Theory Proceedings (ISIT), 2014 IEEE International Symposium on*, 2014.
- [8] C. Rush, A. Greig, and R. Venkataramanan, "Capacity-achieving sparse superposition codes via approximate message passing decoding," *arXiv preprint arXiv:1501.05892*, 2015.
- [9] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, 2011.
- [10] A. Javanmard and A. Montanari, "State evolution for general approximate message passing algorithms, with applications to spatial coupling," *Information and Inference*, 2013.
- [11] J. Barbier, C. Schülke, and F. Krzakala, "Approximate message-passing with spatially coupled structured operators, with applications to compressed sensing and sparse superposition codes," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, no. 5, 2015.
- [12] J. Barbier and F. Krzakala, "Approximate message-passing decoder and capacity-achieving sparse superposition codes," *arXiv preprint arXiv:1503.08040*, 2015.
- [13] J. Barbier, M. Dia, and N. Macris, "Proof of Threshold Saturation for Spatially Coupled Sparse Superposition Codes," in *Information Theory Proceedings (ISIT), 2016 IEEE International Symposium on*, Jul. 2016.
- [14] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *arXiv preprint arXiv:1010.5141*, 2012.
- [15] C. Condo and W. J. Gross, "Sparse superposition codes: A practical approach," in *Signal Processing Systems (SiPS), 2015 IEEE Workshop on*, Oct 2015, pp. 1–6.
- [16] J. Barbier, M. Dia, N. Macris, and F. Krzakala, "The Mutual Information in Random Linear Estimation," in *in the 54th Annual Allerton Conference on Communication, Control, and Computing*, September 2016.
- [17] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with gaussian matrices is exact," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016.
- [18] S. Kudekar, T. Richardson, and R. L. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Trans. on Information Theory*, vol. 59, no. 12, pp. 7761–7813, Dec 2013.

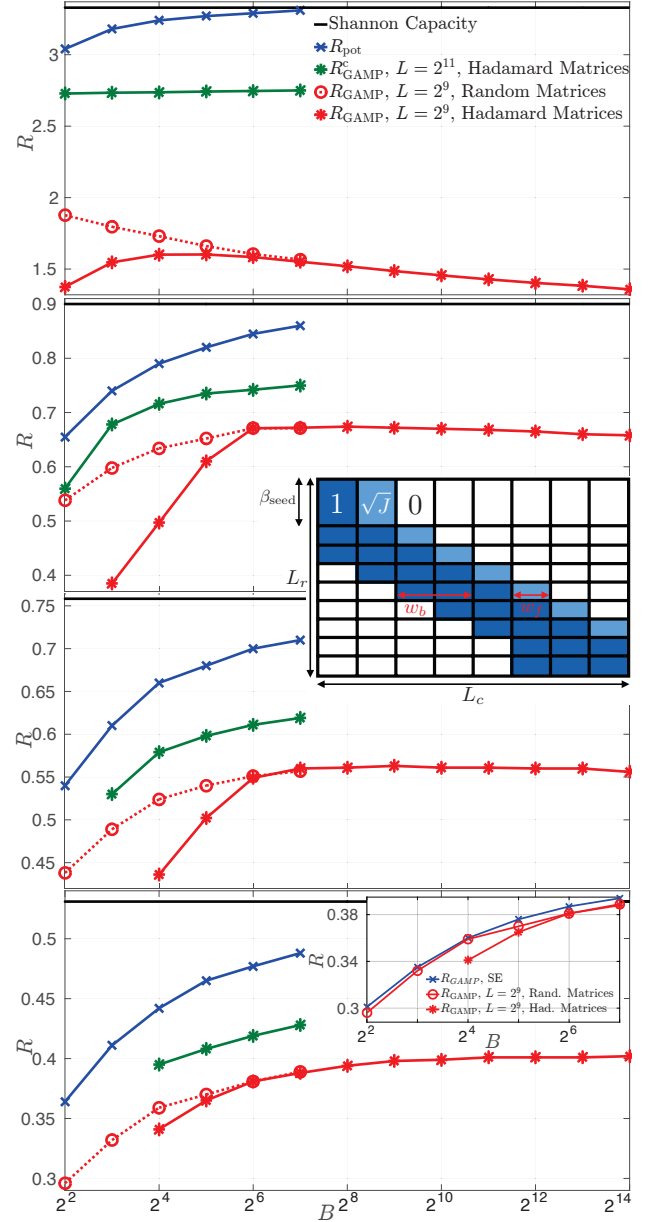


Fig. 3: Phase diagrams for (from top) the AWGNC with $\text{snr} = 100$, the BEC, ZC and BSC all with $\epsilon = 0.1$. The $L \rightarrow \infty$ transition R_{pot} is obtained from the potential by equating its two minima. R_{GAMP} , formally defined for $L \rightarrow \infty$, is instead obtained for finite $L = 2^9$ by running GAMP over 100 instances for each (R, B) and by defining the transition as the highest rate for which at least 50 instances were decoded (up to a small SER due to finite size effects). The bottom inner figure illustrates finite size effects by comparing R_{GAMP} computed in this way (for the BSC) and the “true” $L \rightarrow \infty$ curve predicted by SE; the finite L transition follows very closely the asymptotic one. The two R_{GAMP} curves (dashed and solid red) illustrate that, despite the mismatch in the rates between the Hadamard-based coding matrices and the Gaussian ones for low B , both rates coincide for large B . The region between the red and blue curves is the hard phase. To find the spatially coupled threshold R_{GAMP}^c (green curve), we follow the same procedure as for R_{GAMP} but using coupled Hadamard-based operators and $L = 2^{11}$. These are constructed as described in [11, 12], with the following coupling parameters, see middle inner figure for the block decomposition of a coupled coding operator: number of block-columns $L_c \in \{8, 16, 32\}$; number of block-rows $L_r = L_c + 1$; backward and forward coupling windows $w_b \in \{2, 3, 5, 7\}$, $w_f \in \{1, 2\}$; coupling strength $\sqrt{J} \in [0.53, 0.73]$ for the AWGNC, 0.3 for the other channels (all the blocks other than the light blue coupling ones and the all-zeros blocks have unit strength); relative size of the “seed” block $\beta_{\text{seed}} \in [1.02, 1.25]$.